DOCUMENT RESUME

ED 190 221 PS 011 559

AUTHOR Arias, Beatriz; And Others
TITLE An Evaluation of the Head Start Bilingual Biclutural Curriculum Development Project. Review and Recommendations for the Test Battery.
INSTITUTION Juarez and Associates, Inc., Los Angeles, Calif.
SPONS AGENCY Administration for Children, Youth, and Families (DHEW), Washington, D.C.
PUB DATE Jul 78
CONTRACT 105-77-1048
NOTE 95p.; For related documents, see PS 011 556-561, and PS 011 583. Parts may not reproduce clearly. Best copy available.

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Bilingual Education; Comparative Analysis; Comprehension; Concept Formation; *Curriculum Evaluation; Motivation; Multicultural Education; *Preschool Curriculum; Preschool Education; *Program Evaluation; Self Esteem; Spanish Speaking; *Testing; Tests; *Test Selection
IDENTIFIERS *Project Head Start

ABSTRACT

The purpose of this part of the multimethod Head Start bilingual/bicultural curriculum evaluation (Juarez and Associates 1978), is to familiarize curriculum model developers with test and item analysis procedures used to select and recommend tests for evaluating curriculum models. Considerations and procedures (such as screening, and comparing and determining best fit between tests and programs) used in choosing tests are indicated. Recommendations for measures to be included in the Head Start curriculum model assessment battery are made and test administration guides are provided. Materials related to the test selection effort, including lists of the curriculum models' characteristics, cross-model behaviors, sources of information, a list of language measures, test match forms, results of item analysis, and reviews of recommended and non-recommended measures are appended. (Author/RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED190221

Juárez and Associates, Inc.

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

AN EVALUATION OF THE HEAD START BILINGUAL BICULTURAL CURRICULUM DEVELOPMENT PROJECT

Review and Recommendations for the Test Battery

Authors: Beatríz Arias, Patricia Gándara,
Luis C. Moll, Marlene Zepeda

BEST COPY AVAILABLE

July, 1978

Contract No. HEW 105-77-1048
Project Director: Luis C. Moll, Ph.D.
Prepared for: Dr. Steve Martinez
Project Officer
Dennis DeLoria,
Acting Project Officer
Administration for Children, Youth, and Families
Office of Human Development Services
Department of Health, Education, and Welfare

PS 011559

1100 Glendon Avenue • Los Angeles, California 90024 • (213) 478-0826 • (213) 879-5899

Acknowledgments

A project of this scope and complexity requires the technical expertise and creativity of many individuals. We would like to acknowledge the contributions of the following people: Mr. Rudy Irizzary and Dr. Clint Walker of the Center For the Study of Evaluation, Graduate School of Education, UCLA, who provided valuable technical assistance; Ms. Sara Zapata who helped prepare the report; and Ms. Margaret Mason, who gave editorial guidance.

TABLE OF CONTENTS

I. Introduction 1
Nature of the Evaluation 1
Consideration in Choosing Tests 4

II. Procedures Used in Choosing Tests 8
Selection of Objectives to be Assessed 8
The Search for Instruments 9
Screening the Instruments 10
Language Assessment 13
Language Dominance 14
Spanish Language Comprehension 17
English Language Comprehension 17
Spanish Language Production 18
English Language Production 18
Concept Development 19
Socioemotional Development: Self Esteem and Motivation 22
Determining the Best Fit Between Tests and Programs 26
Recommendations 30
Summary 35

III. Test Admininstration 38
Time of Admininstration 38
Order of Testing 39
Assembling the Students 39
Scoring of the Test 39
Data Analysis 39

Appendixes 42
Appendix A - The model's characteristics and objectives . 40
Appendix B - Cross-model behaviors 44
Appendix C - Sources of information 50
Appendix D - Language Measures 53
Appendix E - Test match forms 57
Appendix F - Results of item analysis 65
Appendix G - Instruments reviewed but not recommended as child impact measures 76
Appendix H - Review of recommended instruments 80

References 87

LIST OF TABLES

Table 1. Test Analysis Criteria 5

2. Rating Form 11

3. Review of English Language Measures 15

4. Review of Spanish Language Measures 16

5. Review of Concept Development Measures 20

6. Review of Self-Concept Measures 23

7. Test Measure and Testing Sequence 36

8. Summary of Recommended Test Battery 37

INTRODUCTION

Among the most difficult aspects of any bilingual evaluation is the selection of appropriate tests or instruments that will provide valid and reliable measures of pupil growth. It has been Juárez and Associates' position that to ensure the selection of appropriate tests (whether norm or criteria-referenced), the curriculum contractors should participate in the selection process. As a consequence, the purpose of this document is to familiarize the curriculum model developers and other interested parties with the procedures utilized to perform the test and item analyses that were used to select and, finally, recommend tests. The methods that are described below are intended to help in selecting tests for use as child impact measures that are of acceptable quality, relevant to the programs, and fair to the students and the teachers.

Nature of the Evaluation

The work statement for the evaluation of the Head Start Bilingual Bicultural Curriculum Development Project (HEW-105-77-1048) specifies that the purpose of the study is to assess the effectiveness of four childhood bilingual/bicultural curriculum models currently under development as part of the Administration for Children, Youth and Families' (ACYF) Head Start Strategy for Spanish-Speaking Children. In particular, the evaluation is to examine:

1. the feasibility of successfully implementing the models in more than one setting;

2. the impact of each model on children, parents, and the Head Start teaching staff; and

3. the attitudes and opinions of Head Start staff and parents regarding the philosophy, objectives, and quality of the program.

In addition, pamphlets have to be developed for transmitting the information obtained in the evaluation in a format that is useful for other parents and teachers interested in building their own bilingual/bicultural Head Start programs.

The work statement also mandates that an experimental evaluation design be followed. It calls for a randomized selection of students to treatment and control groups, and for pre- and post-test change measures to be administered.

It is important to remember, however, that this evaluation is to be carried out in four phases over a three and one-half year period. The first three phases will each last for one year and the fourth will last for six months and will be primarily devoted to writing final reports. During Phase I (October 1, 1977, to September 30, 1978), the evaluation contractors have been selecting the instruments to be used to assess impact and program quality acceptance. They have also made final recommendations on the sites where the evaluation will be carried out and will collect some initial pre-test data from Head Start teachers. During Phase II (October 1, 1978, to September 30, 1979) the evaluation contractor will pilot test the assessment instruments selected during Phase I. Actual data collection will take place during Phase III (October 1, 1979, to September 30, 1980).

Some of the instruments selected by the evaluators to assess program impact will address the performance of Head Start and control children on selected measures of social competence. This will include the following domains:

Spanish language comprehension
Spanish language production
English language comprehension
English language production
Concept development
Socioemotional behaviors

As we have previously stated, it seems that the tendency in evaluations is often to include as many objectives as possible in order to get an idea of what does or does not work. Unfortunately, this practice has often led to spurious findings, primarily owing to a reliance on a single test or task to measure gains. We have suggested and recommended to the curriculum models' staff that we:

1. jointly identify the key objectives under each domain that could become the evaluation's dependent variables;
2. carefully, and again jointly, operationalize these objectives;
3. limit the number of dependent variables to be assessed; and
4. not rely on a single task or test to measure variables, but instead rely on a series of perspectives to look at chosen behaviors in differing contexts and situations. For this we can use a combination of experimental, quasi-experimental, and natural contexts approaches, and employ a variety of techniques to determine what behaviors the subject under study demonstrates (Juárez and Associates, 1978; Moll et al., 1977).

Hence, it is assumed that the psychometric approach must be supplemented by observational and interview strategies to achieve the goals of the evaluation and to enhance the prospects for interpreting measured change in the evaluation's results.

Considerations in Choosing Tests

The purpose of this section is to discuss the guidelines and the procedures that have been followed in making decisions on the measures for this evaluation. After discussing some issues regarding test selection, we describe our procedures in searching, evaluating, and, finally, choosing a test.

Our position is that the most important question to be addressed with respect to testing for program evaluation is: Does the test measure what the program is attempting to do? For this reason, the actual selection of instruments will be guided by the principle of finding the best fit between program and tests. The procedures principally involve comparing tests for technical adequacy and other qualities, and then determining the degree of congruence of the selected tests to the curricular objectives to be evaluated.

Initially, test analysis criteria were generated (see Table 1) from the specific needs of the evaluation and from the criteria previously developed by the Center for the Study of Evaluation (CSE) of the University of California, Los Angeles, to select tests to evaluate educational programs. (Dotseth, Hunter & Walker, 1978; Hoepfner et al., 1974, 1976).

The general idea was that the more comprehensive the criteria used for selecting tests, the more likely we would be to select the best tests available. It became clear, however, that not all of the test features were of equal importance for the purpose of the evaluation. Therefore, using materials made available by CSE,

Table 1

Test Analysis Criteria

The following test analysis criteria evolved from the specific needs of the Bilingual/Bicultural Head Start Evaluation and were influenced in format by criteria previously developed through CSE.

I. Measurement Validity

1. Item Selection -- refers to how effectively the test items are described and justified

2. Face Validity -- refers to how well the test measures specific goal behaviors as determined by a panel of experts

3. Construct Validity -- refers to the relationship of test items to an underlying construct. In other words, does the test measure what it purports to measure?

4. Concurrent Validity -- refers to how well a particular test correlates with another well reputed test

5. Predictive Validity -- refers to how predictive a particular test is in reference to another subsequent behavioral criterion

6. Content Validity -- refers to how closely a test correlates to a specific curriculum

II. Reliability

1. Test-retest Reliability -- refers to how well a test relates to individual repeated trials over time

2. Internal Consistency -- refers to how coherently or consistently the test measures a given behavioral dimension

III. Is Test Designed for Target Population?

1. Utilization by Hispanic -- what particular ethnic groups have previously utilized this test?

2. Utilization by Other Programs/Evaluations -- what other programs or evaluation projects have used this test?

3. Geographical Location -- what parts of the country have utilized this test?

4. Age Group Normed On?

5. Translation based on which ethnic group?

6. Pilot Tested -- on what groups has test been piloted?

IV. Test Format

1. Visual/Auditory Attractiveness -- would preschoolers be attracted to this test instrument?

2. Timing/Pacing -- is it appropriate for preschoolers?

3. Level of Comprehension -- how appropriate is the test's content for preschoolers? (This includes concepts, syntax, and vocabulary of instructions.)

Table 1 (Continued)

Test Analysis Criteria

V. Feasibility of Administration

1. Size of test group
2. Administration -- refers to the quantity of prerequisite training required in order to administer the test
3. Administration Time
4. Ease of Scoring -- how simple is the scoring procedure?
5. Score Interpretation -- how are scores reported/interpreted (frequencies, norms, percentiles, etc.)?
6. Cost

we selected the three most important characteristics a test should have for use in the evaluation.

These three characteristics can be briefly summarized as follows:

1. Efficiency - This refers to procedures for test administration, level of comprehension of the test, the time it takes to be administered, cost, availability, etc. In short, instruments have been examined to determine if they are appropriate and practical to administer. Although the overall time required to take the test was a critical criterion given our target population (preschool children), the tests general suitability for administration by paraprofessional testers hired from the community was another important consideration.

2. Fairness - This is probably the most important criterion, since it not only refers to whether the test was designed for our specific target population, but also refers to the degree to which the test fits or matches the several curricula. Since we consider this congruence characteristic or property of the test vital to the evaluation, our major recommendation is that the curriculum developers also examine and rate each test or subtest recommended for use. In choosing a multilevel test battery, it is our opinion that each separate level of the tests should also be rated by the curriculum specialists who are intimately familiar with the program as it is actually taught. Therefore, as a necessary check on the accuracy of our test ratings, we would like to involve the model developers in the procedures of analyzing the test for congruence. The procedures described in the next section are intended to facilitate this involvement.

3. Technical adequacy - This refers to the validity and reliability of tests as reported by the test-maker, as well as any other important technical characteristics such as excellence of norms, methods of construction, etc. From our discussion of the two other more pragmatic criteria, it should be clear that a test may have good face validity, high reliability, good norms--in short, technical adequacy --but if the objectives which it "measures" are not related to the curriculum objectives being evaluated, the test (or tests) should be considered inappropriate, because it does not include the skills which are covered by the programs.

One additional point may be mentioned at this time; the role of field testing in choosing a test package. It is desirable to try out a test in schools being evaluated or in acceptable similar programs before deciding to use it for evaluation. The teacher's and the students' reactions to the test are a good indicator of its quality and usability (Irizarry, personal communication). As stated earlier, pilot testing of all the instruments will be a major task during Phase II of the evaluation.

Procedures Used in Choosing Tests

The purpose of this section is to present the procedures and the results of the test and item analysis we have performed for each of the separate components under each domain to be evaluated. First the test ratings will be presented, followed by the item analysis to determine program-test congruence.

Selection of Objectives To Be Assessed

We began the identification of objectives by conceptualizing each of the domains for children specified in the evaluation contract. We then reviewed all of the materials available on the four curriculum models to be evaluated in order to describe their characteristics and objectives (see Appendix A). From this extensive list of objectives we inductively generated the domains and relevant behaviors to be assessed in accordance with those specified in the evaluation contract. The result was the identification of several specific curriculum-related components under each evaluation domain.

For example, a domain (e.g., concept development) was first reviewed theoretically, resulting in its tentative breakdown into separate components (e.g., seriation, classification, etc.). These components were then assessed against the four separate curriculum models, through an analysis of the objectives of the various programs to determine if, in fact, these components were going to be taught in the programs. Once these components have been identified, they are then divided into several participant behaviors that can be observed empirically (e.g., the child arranges objects in sequence). At this point we attempted to limit the number of domains and behaviors to be sampled, following our earlier recommendations.

This reduced list of components and their related participant behaviors were then reviewed by the curriculum developers for accuracy. Their suggestions were incorporated into a revised list. At the conclusions of these initial procedures we had a manageable and accurate list of across-model objectives to be assessed (see Appendix B).

The Search for Instruments

A search was conducted in order to locate and review appropriate instruments important to each of the specified impact domains. Instrument selection began with a comprehensive review of other national evaluations of early childhood programs with special emphasis on programs that focused on bilingual/bicultural population. Documents reviewed include Shipman's Head Start Longitudinal Study (1972); Walker et al.'s review of measures used in Head Start Planned Variation (1973); Deloria et al.'s National Home Start Evaluation (1974); Love

et al.'s Project Development Continuity (1975); Raizen and Bobrow's Design for a National Evaluation of Social Competence in Head Start Children; Stallings et al.'s Phase II Instruments for the National Day Care Cost-Effects Study; and the American Institute for Research's Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program.

In addition, a comprehensive search was made of the ERIC system, ETS Test Collections, and the resources of the Bilingual/Bicultural Dissemination Center. We also conducted a major mail-out to commercial publishers of tests for young children. Every attempt was made to review instruments suggested by the curriculum contractors, consultants to our project, child development researchers, and practitioners. Appendix C contains a list of the sources of information utilized in the search for appropriate instrumentation.

Screening the Instruments

After generating a large pool of tests, we began screening them according to the needs of the evaluation and the purpose the tests are to serve (see Table 2). Among the criteria used to reduce the original list of tests to manageable size were 1) technical quality, 2) the test's appropriateness for use with preschool-aged Hispanic children, and 3) practicality, or, for example, the availability of the test in Spanish, whether individual administration was possible, the time it took to administer the test, feasible cost, etc. (Walker, 1978). Tests that failed to meet these and similar criteria were considered unacceptable. These screening procedures allowed us to identify the preferable tests for the purpose of the evaluation.

TABLE 2

RATING FORM

NAME OF TEST _____ PUBLICATION DATES _____

SUBTEST DESCRIPTION _____

I. Measurement Validity	High	Medium	Low	Comments
1. Item Selection	described in detail	poorly described or justified	no description provided	
2. Face Validity	very strong convincing	moderate evidence	weak, sketchy evidence	
3. Construct Validity	very strong convincing	moderate evidence	weak, sketchy evidence	
4. Concurrent Validity	r > .70	.70 > r > .30	none reported; r less than .30	
5. Predictive Validity	r > .70	.70 > r > .30	not reported	
6. Evaluation Validity	highly correlated	moderately correlated	no correlation	
II. Reliability				
1. Test-retest	r=.70 to .90	r=less than .70	r not reported	
2. Internal Consistency	r=.70 to .90	r=less than .70	r not reported	
III. Is Test Designed for Target Population?				
1. Utilization by Hispanics				specify groups
2. Utilization by other programs/evaluations				specify
3. Geographical Location				specify
4. Age Group Normed On?				specify
5. Translations based on which groups?				specify groups
6. Pilot tested on which groups?				specify groups
IV. Test Format				
1. Visual/Auditory Attractiveness	high quality realistic	moderately attractive	not good quality	
2. Timing/Pacing	highly age appropriate	moderately age approp.	not age appropriate	
3. Level of Comprehension	highly age appropriate	moderately age approp	not age appropriate	
V. Feasibility of Administration				
1. Size of Test Group				specify
2. Administrator	participant researcher	school personnel	specialist	

	High	Medium	Low	Comments
3. Administration Time	20 min or less	30 min or less	40 min or less	
4. Ease of Scoring	highly obj. and simple	moderately obj. and simple	subjective; difficult	
5. Score Interpretation	straight forward	moderately difficult	novel, ambiguous; non-existent	
6. Cost	low	moderate	high	

RECOMMENDATION:

Language Assessments. Of particular importance to the evaluation effort is the assessment of the level and depth of the children's language competence on entry to the programs. This is not only important for determining growth or outcome at the conclusion of the evaluation, but it is essential to consider language abilities as intervening variables which may mediate cognitive or academic outcomes. Phrased in question form, given the entry-level language abilities of the children, what form of instructional treatments promote beneficial cognitive and academic outcomes? It should, therefore, be clear that in addition to the tests mentioned in this document, which will be administered in a pre-post fashion, teacher and parental opinions, and classroom observations will also be used to assess language in a variety of situations and contexts and as a continuous process throughout the academic year.

In addition, the issue of assessing language abilities among children of very different backgrounds and characteristics must be carefully addressed. The heterogeneity of the Hispanic children who will participate in the programs should be obvious. There will be significant differences--regional, rural/urban, historical, linguistic, ethnic, racial--among children. For example, the Los Angeles site will be urban, with most of the children possibly being second generation; the New Mexico site is rural with most of the children also being second generation; a Texas site is located in a small town near the border, with many of the children being first-generation, monolingual Spanish speakers; and the New York City sites are urban with a variety of children from the Caribbean and Central and South America, and Black Americans. The variations

between sites, each with unique and special characteristics, place special constraints on both the test selections and the interpretation of results (Love et al., 1975).

The search for language instruments produced a large quatity of tests (see Appendix D) that had to be screened further, primarily for their appropriateness for use with preschool populations. These exclusionary criteria reduced the many tests that were considered to a more manageable number. The remaining tests, those that "survived" the screening, were then compared to one another and judged for their appropriateness (see Tables 3 and 4). We concluded that the most suitable language measures are as follows:

Language Dominance

- *Bilingual Syntax Measure* (Harcourt, Brace and Jovanovich, Inc.)

The BSM is designed to measure children's oral proficiency in English and/or Spanish grammatical structures by using natural speech as a basis for making judgements. A comparison of proficiency levels in two languages yields information that indicates the dominant language and degree of bilingual proficiency.

- *Del Rio Language Screening Test* (National Educational Laboratory Publishers, Inc.)

The DRLST is designed to identify difficulties in the child's Spanish and English language abilities. In addition, it may be used as a test of relative bilingual proficiency. The DRLST consists of 5 subtests of receptive-vocabulary, sentence-repetition-length, sentence-repetition-complexity, oral commands and story comprehension.

- *Pictorial Test of Bilingualism and Language Dominance* (South Texas Testing Service)

This test assesses language facility in English and Spanish. Results from test place children in six diagnostic categories: 1) English dominant, 2) Spanish dominant, 3) bilingual, 4) language deficient, 5) needs further testing, 6) pseudo-bilingual. Part I, word recognition, Part II, production.

Table 3

Review of English Language Measures

Criteria	Comprehension					Production		
	LTS	HWS	Del Rio	PTBLD	TBCE/S	YSI	BSM	PTBLD
Measurement								
Item Selection	+	+	0	0	0	0	*	0
Face Validity	+	+	+	+	+	+	*	+
Construct Validity	0	0	0	0	0	+	*	+
Concurrent Validity	0	0	0	+	0	+	*	0
Predictive Validity	0	0	0	0	0	+	*	0
Content Validity	+	+	–	–	–	+	+	–
Reliability								
Test-retest	0	0	.86-.88	.35-.96	0	0	*	.35-.96
Internal	.77	.68-.92	0	.85,.89**	0	.69-.89	*	.85,.89**
Appropriateness to Population								
Utilization by Hispanics	0	0	–	–	0	+	+	–
Utilization by Programs/ Evaluations	0	0	–	0	0	–	+	–
Geographic Location	–	–	–	–	0	+	+	–
Age Group Normed On	+	+	+	+	0	+	+	+
Translation	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Pilot Tested	–	–	+	+	+	–	+	–
Test Format								
Visual/Auditory	+	+	+	+	+	+	+	+
Timing/Pacing	+	+	+	+	+	+	+	+
Level of Comprehension	+	+	+	+	+	+	+	+
Feasibility of Administration								
Size of Test Group	+	+	+	+	+	+	+	+
Administration Time	+	+	+	+	+	+	+	+
Ease of Scoring	+	+	–	+	+	+	+	+
Score Interpretation	+	+	+	+	+	+	+	+
Cost	+	+	–	+	+	+	+	+

Note: LTS=Circus subtest--Listen to the Story
HWS=Circus subtest--How Words Sound
Del Rio=Del Rio Language Screening Test
PTBLD=Pictorial Test of Bilingualism and Language Dominance
TBCE/S=Test of Basic Competence in English and Spanish
YSI=El Circo subtest--You Say It
BSM=Bilingual Syntax Measure
** = .85 (English), .89 (Spanish)

Legend: + = instrument satisfies criterion
– = instrument doe not satisfy criterion
0 = information not available
N/A = not applicable
* = technical information not available at time of writing

Table 4

Review of Spanish Language Measures

	Comprehension						Production		
Criteria	Del Rio	EEC	CSLP	PTBLD	SWCEL	TBCE/S	BSM	DT	PTBLD
Measurement									
Item Selection	0	0	0	0	0	0	*	0	0
Face Validity	+	+	+	+	+	0	+	+	+
Construct Validity	0	0	0	0	0	0	*	0	0
Concurrent Validity	0	0	0	-	0	0	*	0	+
Predictive Validity	0	0	0	0	0	0	*	0	0
Content Validity	-	+	+	-	-	-	*	+	-
Reliability									
Test-retest	.66-.88	0	*	.35-.96	.91	0	*	0	.35-.96
Internal	0	*	*	.85,.89**	0	0	*	****	.85,.89**
Appropriateness to Population									
Utilization by Hispanics	-	+	+	-	-	0	+	+	-
Utilization by Programs/ Evaluations	-	-	-	-	-	0	+	-	-
Geographic Location	-	+	+	-	-	0	+	+	-
Age Group Normed On	+	+	+	+	+	0	+	+	+
Translation	-	+	+	-	-	0	+	+	-
Pilot Tested	+	+	+	+	+	+	+	+	-
Test Format									
Visual/Auditory	+	+	+	+	+	+	+	+	+
Timing/Pacing	+	+	+	+	+	+	+	+	+
Level of Comprehension	+	+	+	+	+	+	+	+	+
Feasibility of Administration									
Size of Test Group	+	+	+	+	+	+	+	+	+
Administration	+	+	+	+	+	+	+	+	+
Administration Time	+	+	+	+	-	+	+	+	+
Ease of Scoring	+	+	+	+	+	0	+	+	+
Score Interpretation	-	+	+	+	+	+	+	+	+
Cost	+	+	+	+	-	+	+	+	+

Note: Del Rio=Del Rio Language Screening Test
EEC=El Circo subtest--Escuchen Este Cuento
CSLP=El Circo subtest--Como Suenan Las Palabras
PTBLD=Pictorial Test of Bilingualism and Language Dominance
SWCEL=SWCEL Comparative Language Dominance Test
TBCE/S=Test of Basic Competence in English and Spanish
BSM=Bilingual Syntax Measure
DT=El Circo subtest--Dimelo Tu
** = .85 (English), .89 (Spanish)
*** = .71-.79 (I,III),.25-.48 (II)

Legend: + = instrument satisfies criterion
- = instrument doe not satisfy criterion
0 = information not available
N/A = not applicable
* = technical information not available at time of writing

- *SWCEL Comparative Language Dominance Test* (Southwestern Cooperative Education Laboratory, Inc.)

This test is designed to measure English and Spanish language comprehension and language dominance. Measures Spanish of the New Mexico/ Southern Colorado vernacular. Yields scores for receptive vocabulary.

Spanish Language Comprehension

- *El Circo subtest - Escuchen Este Cuento y Como Suenan Las Palabras* (Educational Testing Service)

Escuchen Este Cuento y Como Suenan Las Palabras are tests designed to tap language comprehension, interpretation, functional language use, vocabulary and recall of oral language.

Como Suenan Las Palabras - auditory discrimination -- discriminating among phonemes, including initial and final consonants and medial vowels.

- *Del Rio Language Screening Test* (see page 10)
- *Pictorial Test of Bilingualism and Language Dominance* (see page 10)
- *SWCEL Comparative Language Dominance Test* (see above)
- *Test of Basic Competence in English and Spanish* (Cervenka)

This test is designed to measure English and/or Spanish language proficiency. This is an experimental test designed to measure basic language competence of children who speak English or Spanish or who are bilingual. Normed on Texas children, it was developed for three through six year olds attending preschool programs. Nine subtests assess oral vocabulary, comprehension of commands, recognition of interrogative patterns, phonemic discrimination at word level, production of grammatical directions, assimiliation of meaning and phonemic discrimination at sentence level. Technical manual not available.

English Language Comprehension

- *Circus subtest - Listen to the Story and How Words Sound* (Educational Testing Service)

This test is designed to tap comprehension, interpretation and recall of oral language.

- *Del Rio Language Screening Test* (see page 10)
- *Pictorial Test of Bilingualism and Language Dominance* (see page 10)
- *Test of Basic Competence in English and Spanish* (see page 11)

Spanish Language Production

- *El Circo subtest - Dímelo Tú* (Educational Testing Service)

  There is a Spanish version with a corresponding English version. Productive language: describing common objects, using functional language, and telling a story based on a picture (3 parts).

- *Bilingual Syntax Measure* (see page 10)
- *Pictorial Test of Bilingualism and Language Dominance* (Part 2) (see page 10)

English Language Production

- *El Circo - You Say It* (Educational Testing Service)

  This is the English version (see above)

- *Bilingual Syntax Measure* (see page 10)
- *Pictorial Test of Bilingualism and Language Dominance* (Part 2) (see page 10)

After carefully comparing one test to another, and their relative advantages and disadvantages, the following tests were subjected to an item analysis (details are provided in the next section) in order to determine if they adequately matched the curriculum objectives.

- Bilingual Syntax Measure
- Del Rio Language Screening Test

- CIRCO subtests - Escuchen Este Cuento, Como Suenan Las Palabras, Dímelo Tú, and You Say It
- CIRCUS - Listen to the Story and How Words Sound

Concept Development. In order to meet the requirements of the evaluation contract to measure concept development, the Juárez and Associates team did a thorough search of current instruments available in both English and Spanish. Few tests which are appropriate for use with preschool age children have been developed with Hispanic children in mind. Even fewer offer versions of the tests in both languages.

It should be noted that there is no requirement that a test have parallel forms in both languages. However, it is important that the types of behaviors which are sampled by the Spanish and English test forms be substantially similar. That is, English speaking children should be required to produce similar behaviors as Spanish speakers. For this reason one criterion for selection was that a test had to have both an English and a Spanish form whose psychometric properties had been tested on the appropriate populations. From the small pool of tests which met all of our criteria (also see Table 5), we selected several of the Circo/Circus subtests (Pensándolo Bien/Think it Through; Mira y Recuerda/See and Remember; Cuánto y Cuántos/How Much and How Many; Parecidos/Look-Alikes), the Preschool Inventory (PSI), the Walker Readiness Test, and the TOBE 2.

In comparing these tests, the Walker test was rejected because

Table 5

Review of Concept Development Measures

Criteria	English					Spanish				
	How Much & How Many	Look Alikes	PSI	TOBE 2	Walker	Cuanto & Cuantos	Parecidos	PSI	TOBE 2	Walker
Measurement										
Item Selection	+	+	+	*	+					
Face Validity	+	+	+	-	+	+	+	+	*	+
Construct Validity	+	+	+	*	+	+	+	+	-	+
Concurrent Validity	0	0	.90**	*	+	0	0	0	*	0
Predictive Validity	0	0	+	*	+	0	0	0	*	0
Content Validity	-	-	+	-	-	0	0	0	*	0
						+	-	+	-	-
Reliability										
Test-retest	0	0	.83	*	.86-.91	0	0	0	*	0
Internal	.61-.87	.84	.83	*	.87-.89	>.70	.72	0	*	0
Appropriateness to Population										
Utilization by Hispanics	0	0	+	0	+	+	+	+	0	+
Utilization by Programs/ Evaluations	0	0	+	+	0	0	0	+	-	0
Geographic Location	+	+	+	+	+	+	+	+	-	-
Age Group Normed On	-	-	+	+	+	+	+	+	0	+
Translation	N/A	N/A	+	N/A	N/A	+	+	+	+	-
Pilot Tested	-	-	+	+	+	+	+	+	-	-
Test Format										
Visual/Auditory	+	+	+	+	+	+	+	+	+	+
Timing/Pacing	+	+	+	+	+	+	+	+	+	+
Level of Comprehension	+	+	+	-	-	+	+	+	-	-
Feasibility of Administration										
Size of Test Group	+	+	+	+	+	+	+	+	+	+
Administration	+	+	+	+	+	+	+	+	+	+
Administration Time	-	+	+	+	+	+	+	+	+	+
Ease of Scoring	+	+	+	+	+	+	+	+	+	+
Score Interpretation	+	+	+	+	+	+	+	+	-	+
Cost	+	+	+	+	+	+	+	+	+	+

Note: PSI=Preschool Inventory
TOBE=Test of Basic Experiences
Walker=Walker Readiness Test
**= >.90 with 64 item version

Legend: + = instrument satisfies criterion
- = instrument does not satisfy criterion
0 = information not available
N/A = not applicable
* = technical information not available at time of writing

it samplies too narrow a range of behaviors, has poor instructions in both languages, and did not fit very will with model curricula (see next section). The TOBE 2 was rejected on the basis of format, item content, and failure of the publisher to provide information regarding item development and standardization for Hispanic samples. The Pensándolo Bién/Think it Through subtest of the Circo was rejected on the basis of format (poor graphics) and confusing content. This phase of the screening left us with a conbination of Circo/Circus subtests (Mira y Recuerda/See and Remember; Cuánto y Cuántos/How Much and How Many; Parecidos/Look Alikes) versus the PSI. These tests, which were then examined for their program congruence, are briefly described below.

CIRCO — *CIRCUS*

Mira y Recuerda — *See and Remember*

Visual and associative memory: remembering pictures, name associations -- both immediately and after intervening experiences (20 items).

CIRCO — *CIRCUS*

Cuanto y Cuantos — *How Much and How Many*

Quantitative concepts: demonstrating understanding of enumeration, counting, one-to-one correspondence, ordination, comparison, quantitative language, through the identification of appropriate pictures (40 items).

CIRCO — *CIRCUS*

Parecidos — *Look-alikes*

Visual discrimination: matching identical letters, numbers, or drawings (27 items).

Preschool Inventory

This test assesses general cognitive skills and concepts. It taps a range of verbal, quantitative and perceptual motor skills.

Socioemotional Development: Self-Esteem and Motivation. The inclusion of self-esteem and motivation as direct outcome variables for the Bilingual/Bicultural Head Start Evaluation reflects the recognition by ACYF of the important role of affect in the educational process. However, the measurement of these constructs presents a variety of problems and constraints that are intensified when the focus is young children.

The first problem confronted concerns the definition of self-esteem and motivation. As noted by Coller (1971), the conceptualization of self-concept includes a wide array of behavioral components which encompass self-esteem and motivation. This being the case, self-esteem and motivation can be operationalized as constituents of self-concept (see Table 6). Consequently, as an initial strategy, standardized tests falling under this general rubric were investigated. Basically, this procedure proved to be unfruitful. It was problematic not only because of the inadequacies of the available instruments, but also because the construct self-concept circumscribes a broad configuration of behaviors. Thus, we are in agreement with Stallings, Wilcox, and Travers (1976) that the use of self-concept as an outcome variable is less likely to yield relevant results than other outcome variables, since it is so indirectly tied to specific behaviors. As a consequence, the selection of any one self-concept test cannot exhibit the multiplicity of affective behaviors demonstrated by a young child.

Secondly, the lack of an exact theoretical base raises special problems for the assessment and interpretation of self-concept in

Table 6

Review of Self-Concept Measures

Criteria	Brown IDS	CSCT	ISCT	LSCT	DiLorenzo et al	PSCPT	PSCI	RPD	SCAMin	Butler & Swayze	Self Social	Thomas
Measurement												
Item Selection	0	0	+	-	0	0	+	0	0	0	0	0
Face Validity	+	+	+	+	+	+	-	-	+	+	+	+
Construct Validity	+	-	-	-	-	0	-	-	+	-	-	+
Concurrent Validity	0	0	0	0	0	0	0	0	0	0	0	0
Predictive Validity	-	0	0	0	0	0	-	-	0	0	0	0
Content Validity	-	-	-	-	-	-	-	0	-	-	-	-
Reliability												
Test-retest	.59-.64**	0	0	0	0	.80,.94	.57-.91	0	.79	0	.48-.85	.78
Internal	+	0	.66	0	0	0	0	0	0	0	0	.73
Appropriateness to Population												
Utilization by Hispanics	-	0	0	0	0	0	+	0	+	0	0	0
Utilization by Programs/ Evaluations	+	0	+	0	0	0	0	0	+	0	+	0
Geographic Location	+	0	-	-	-	-	+	0	+	-	-	0
Age Group Normed On	+	+	+	+	+	+	-	0	+	0	+	0
Translation	0	0	0	0	0	0	0	N/A	+	0	0	0
Pilot Tested	-	-	-	-	-	-	+	0	+	0	0	0
Test Format												
Visual/Auditory	+	+	N/A	+	N/A	+	+	N/A	+	N/A	-	+
Timing/Pacing	+	+	N/A	+	N/A	+	+	N/A	+	N/A	+	+
Level of Comprehension	+	+	N/A	+	N/A	+	+	N/A	+	N/A	-	+
Feasibility of Administration												
Size of Test Group	+	+	+	+	N/A	+	+	+	+	+	+	+
Administration	+	0	+	+	N/A	+	+	-	+	+	+	+
Administration Time	+	+	+	+	N/A	+	+	+	+	+	+	+
Ease of Scoring	+	0	+	+	+	+	+	+	+	+	+	+
Score Interpretation	+	0	+	+	+	+	+	-	+	0	+	+
Cost	+	0	+	0	0	0	+	0	+	0	+	+

Note: Brown IDS=Brown Self-Concept Referents Test
CSCT=Children's Self-Concept Test
ISCT=Inferred Self-Concept Judgement Scale
LSCT=Learner Self-Concept Test
DiLorenzo et al= Teacher Measurement of Pupil Self-Concept
PSCPT=Preschool Self-Concept Picture Test
PSCI=Primary Self-Concept Inventory
RPD=Riley Preschool Developmental Screening Inventory
SCAMin=Self-Concept and Motivation Inventory
Butler&Swayze=The Self-Concept Sub-scale of Evaluation
Self-Social=Self-Social Constructs Test
Thomas=Thomas Self-Concept Values Test
** = ETS,Longitudinal Version

Legend: + = instrument satisfies criterion
- = instrument doe not satisfy criterion
0 = information not available
N/A = not applicable

young children in particular. According to Walker (1973), the measurement of socioemotional outcomes in young children must take into consideration their short attention span, the fludity of their feelings, moods, and attitudes across time and situations, and finally their "egocentric" perspective as limiting their ability to take the viewpoint of thers. Moreover, there is some evidence to show that young children are predisposed to select socially desirable attributes during test situations (Shipman, 1972).

The paucity of appropriate instruments and the review of previous evaluation studies lead us to conclude that the assessments of socioemotional development should be done through observations of specifiable behaviors. These behaviors are described in Appendix B.

A variety of observational instruments, check lists, and ratings are currently being considered and will be piloted during Phase II of the evaluation. Findings about particular child variables obtained from these sources of data can be cross-validated with ethnographic observations to see if the different sources obtain similar information.

These instruments include the following:

Child Observation Record (Sections 10, 11, 12, and 13) (High Scope Educational Research Foundation)

This checklist focuses on socioemotional development within the following sub-categories: autonomy and relationship with adults, general level of involvement with peers, social problems solving and empathy and pride in work.

Child Observation System/Prescott (Stanford Research Institute)

This instrument is designed to measure related socioemotional indices such as autonomy, dependency, aggression, social involvemet, prosocial behavior, problem solving, curiosity, and task persistence in 3 and 4 year olds.

Schaefer/Aaronson Behavior Inventory

The teacher rating scale assesses such socioemotional indices as dependence, autonomy, aggression and social involvement, cooperation and social and emotional harm.

All of the stated observation systems will be further evaluated in Phase II.

In a sense, the assessment of the socioemotional behaviors will examine the extent to which the programs are relevant to the cultural backgrounds of the children. If the curriculum is culturally appropriate or embedded in the culture of the participants, then the children should be expected to display the positive bahaviors that compose this socioemotional component. If the child is unable to adjust to the preschool environment, or feels bad or embarrassed about participating, then the curriculum is probably not culturally appropriate.

However, the specific cultural objectives of the programs for the parents and the children vary considerably between models (such objectives would encompass active involvement of parents and community residents, attitudes and values of the teachers, incorporation and reflection of life experiences in the classroom, basic knowledge of culture on the part of the children, and cultural artifacts in the classroom). Consequently, classroom observations, interviews, and ratings will be used to assess model specific components. The model developers are urged to share with the evaluators any new or modified cultural objectives that have been developed in the past year.

Determining the Best Fit Between Tests and Programs

The purpose of the screening was to quickly and efficiently reduce the number of tests under consideration. A crucial next step in choosing the tests was to carefully examine the instruments in more detail for their curricular relevance. The procedures described here are intended to be repeatable by the curriculum developers, since the procedures call for the tests and the curriculum objectives to be examined together.[1]

To carry out this item-objective analysis of congruence, we first used the list of across-model objectives to form a worksheet called the Test Match Form (see Appendix E). The form consists of a series of column where the steps of the process are recorded. The first column consists of the objectives of the instructional program (s). In the second column you list the test items according to how they match the given objectives. In column three you judge the degree of match between each program objective and test item. In column four you judge the degree of item difficulty. This whole process is aided by familiarity with how the objectives are to be taught.

The procedures are outlined below in more detail in order to facilitate replication.

STEP 1 (Column 1) List of the curricular objectives. These are major component and participant behaviors within the domains being tested.

---

[1]Our thanks to Clint Walker and Ruddie Irizarry from the Center for the Study of Evaluation, UCLA, for sharing with us the test analysis procedures they are developing.

Suggestions to the curriculum developers: Since you have a more intimate knowledge of your own programs, pay particular attention to whether there are important objectives missing from our list, or whether the components added to the list are relevant or irrelevant to your curriculum. We assume that there might be additions or modifications to the curriculum materials which we have reviewed. Any relevant updates should be carefully noted, since the major purpose of this test item-curriculum comparison is to find or determine the test battery which comes closest to "measuring" what the program is trying to teach.

STEP 2 (Column 2) List of test items. The purpose is to determine how the items match the program objectives. For each test item you are asked to decide whether it does or does not match a curricular objective.

Suggestions: You may want to do this by matching separately for each test or for the entire test battery. We found it useful to do the matching by evaluation domain first before looking at the composite match of the entire test battery. This made the task more manageable. Also, it is advisable that more than one person judge or do the matching independently, and then share the results for discussion. In addition, a given test item may be assigned or matched to more than one objective, although the judged degree of match may vary (column 3). However, multiple matchings to a variety of objectives may indicate that the test lacks divergent validity, a common problem with standardized tests. The results of this procedure should give a good indication of whether the test battery adequately covers important aspects of what is being taught. Of course, given time

constraints on testing young children, it is highly unlikely that any test battery will cover every objective. This procedure will ensure, however, that the most important objectives, or those objectives which are most likely to be affected by the treatment, will be adequately assessed by the measures employed.

STEP 3 (Column 3) - Judged degree of match. In this column we estimated how the test items match the respective program objectives. If we felt that the degree of match to the objective was poor, we assigned it a rating of zero; if adequate, a rating of 1; if good, a rating of 2.

Suggestions: It is recommended that the degree of match be determined by the format, content, and the student behaviors actually taught. In other words, let us assume that a concept development test purports to assess the children's understanding of time. If the test's task demands are similar to the skills actually taught in the classroom, we might judge the degree of match as adequate or possibly good. But if the test task demands are completely dissimilar to what is actually taught in the classroom, then the degree of match is poor. In making this determination, the curriculum developer's knowledge of what actually goes on in the classroom will facilitate making an accurate judgment of the degree of match.

STEP 4 (Column 4) - Judged degree of item difficulty. In this column you simply judge if the item is too difficult, just right, or too easy for the children who are going to be tested. If the item is too easy or too hard, assign a rating of 0. If it is appropriate, assign a rating of 1; if it is just right, assign a rating of 2.

Suggestions: For the norm-referenced tests, information is usually provided by the test-makers on the percentage of children from their norm group who pass each individual item, range of scores, mean scores by test, and so forth. This information, when available, may prove helpful in assessing degree of item difficulty.

STEP 5 (Column 5) - Comments. In this column any comments, suggestions, or ideas regarding the tests or the test selection procedures can be recorded. After completing the Test Match Form, the number of objectives which are matched appropriately by the test items or the overall number of unmeasured objectives can be tabulated. How thoroughly and appropriately the test matches the important and relevant objectives determines whether the test is recommended for use.

Results of the test congruence analysis for the language and concept development tests using the procedures outlined above may be found in Appendix F. The test match forms of four concept development tests are included so the reader can compare the results. However, only one test will be recommended.

The test battery recommended for use in the evaluation is described in the next section.

Recommendations:

After a thorough search for instruments, rigorous screening, comparison, and selection of the tests, item-by-item analysis of congruence with the curricular objectives, and consultation with several experts in the field, we are ready to recommend a test battery that we are confident is appropriate and fair, and that has a logical relationship to what is being measured.

Language Dominance

Recommended: Bilingual Syntax Measure

Reasons: We originally questioned the usefulness of assessing language dominance, particularly if we are obtaining language proficiency measures. However, it is essential to obtain accurate and empirical entry-level data to cross-validate teacher and parental opinions regarding the bilingual fluency of the children and to determine the sequence to test the children. It is also necessary to verify the criterion specified in the contract that 50 percent of the students must be Spanish dominant or bilingual. Therefore, we decided to select a test that would measure language proficiency as well language dominance. The Bilingual Syntax Measure was chosen over the other dominance tests because of its brief 10-15 minute administration time and because there are no special requirements for the examiner, except that s/he must be bilingual and bicultural. In contrast to most other language dominance tests, the Bilingual Syntax Measure has been field-tested with Mexican American, Anglo, Puerto Rican, and Cuban children. The test has also

been used in other national evaluations and has received positive critical reviews from a number of experts in the field of bilingual assessment.

Spanish Language Comprehension

Recommended: CIRCO subtests -- Como Suenan Las Palabras and Escuchen Este Cuento.

Reasons: The CIRCO subtests were chosen, in part, because the CIRCO battery has been developed specifically for Spanish-speaking 4-, 5-, and 6-year-old children. CIRCO is one of the few test batteries suitable for use with children from Mexican American, Puerto Rican, and Cuban backgrounds. It also has adequate technical quality and matches the curriculum objectives well.

English Language Comprehension

Recommended: How Words Sound and Listen to the Story.

Reasons: According to the test-makers (Cordova, 1977), although the CIRCO and CIRCUS tests resemble each other, the two test batteries are not parallel and cannot be equated. This point is well taken. However, in our view it is important that the task demands or the types of behaviors sampled in assessing the Spanish-speaking students be the same or highly similar to those used to assess English-speaking children. This, in part, led us to recommend these two CIRCUS subtests. Of equal importance was the fact that these two subtests of CIRCUS,in combination, provided the best curriculum match within reasonable time constraints of any tests

reviewed. In addition,Listen to The Story and Escuchen Este Cuento were selected because they tap several items under the Concept Development component and thus strengthen the battery in that area without requiring another test.

It should be noted that the intent in this evaluation is to assess all of the children in both languages; therefore, we encounter the problem of "test wiseness" if the test items of the English and Spanish test are highly similar, as is the case with Escuchen Este Cuento and Listen to the Story. In order to control for test wiseness,for language comprehension the children will be tested initially in their second language. We are assuming that most children will only be able to complete part of the test in their second language, and they will be exposed to a minimum number of items from the test. Then, to reduce the effects of memory, the first language comprehension pretest will be the last test in the battery to be administered. This will place an interval of two to three weeks between the administration of the two tests.

For example, if a child is Spanish dominant or monolingual, the first test s/he will receive from the battery will be the English language comprehension test. Two or three weeks later (the entire battery will be administered in three weeks except for the initial BSM) s/he will receive the Spanish language comprehension test.

If our assumption that most children will only be able to complete part of the test in their second language is erroneous, and most children enter the programs bilingual, then the two-to-three week interval between tests in our only control for test wiseness. The measures and the testing sequence to be used are presented in Table 7.

Spanish Language Production

Recommended: CIRCO subtest--Dímelo Tú

Reasons: This subtest matches the cross-model objectives well. It includes an English version that was specifically designed to assess English production in Spanish-speaking children and that can also be administered to English-speaking Hispanic children. As with the other CIRCO tests, it was piloted and normed on Mexican American, Puerto Rican, and Cuban children.

English Language Production

Recommended: CIRCO subtest -- You Say It

Reasons: This is the corresponding English version of Dímelo Tú mentioned above. There are enough differences in item content that both tests can be administered without encountering the problems of learning or test wiseness discussed above. However, since second language production will also be assessed by the Bilingual Syntax Measure , we recommended that the CIRCO's Dímelo Tú and You Say It only be given in the child's first language.

For example, if a child is Spanish dominant or monolingual, the second language, English, will be assessed by the pre/post use of Bilingual Syntax Measure, while Spanish language production will be tested with CIRCO's Dímelo Tú. In such a case, You Say It will not be used, thus reducing the individual testing time.

Concept Development

Recommended: Preschool Inventory(SRI 32-item version)

Reasons: Because this evaluation attempts to assess linguistic development in two languages, several areas must be assessed twice. The time required to administer a test therefore becomes a critical factor in the selection of instruments. While concept development will only be tested in one -- the dominant -- language, it is essential that it not overburden an already lenghty test battery. It was concluded that a minimum of two subtests of the CIRCO/CIRCUS battery would be required in order to sample a broad enough range of the behaviors which are commonly categorized as "concept development" and which are included in the four curricula to be evaluated. Even limiting this assessment to only two CIRCO/CIRCUS subtests would require approximately 30 minutes of testing time. In an effort to cut back on test administration time, it was decided to recommend the PSI (SRI 32-item version), which samples a range of behaviors similar to those sampled by the CIRCO subtests (see Appendix F), requires only about 15 minutes to administer, and is standardized on 3-, 4-, and 5-year-old Head Start subjects including Spanish speakers for the Spanish-language version.

Although the PSI has been criticized for having a "middle-class bias," the puslishers point to data which demostrates that the test does not discriminate between SES or racial/ethnic group gain scores. It should also be noted that the CIRCO/CIRCUS test scores are equally affected by SES and racial/ethnic differences. An additional factor which

influenced our decision is that the PSI has been demonstrated to be highly sensitive to educational intervension and has been used in a number of large-scale evaluations.

As mentioned earlier, the concept development assessment is further strengthened by items from Listen to the Story/Escuchen Este Cuento. Information processing conceptual inference, understanding of relational terms, and time relationship are aspects of concept development which are included in this subtest.

Summary

On the basis of the test and item analysis described in this document, we recommend the use of the following instruments in the Fall 1979 test battery:

- Bilingual Syntax Measure
- CIRCO subtests: Como Suenan Las Palabras
  Escuchen Este Cuento
- CIRCUS subtests: How Words Sound
  Listen to the Story
- CIRCO subtest: Dímelo Tú
  You Say It
- Preschool Inventory: (SRI 32-item version)

Table 7 summarizes the measures and the testing sequence recommended.

Table 8 summarizes the child variables and the instruments expected to assess the variables.

Table 7

Test Measures and Testing Sequence

	Spanish Dominant		English Dominant	
	(1st) Spanish	(2nd) English	(1st) English	(2nd) Spanish
Week 2 Dominance/ 2nd Language Production	BSM* (15 min)		BSM* (15 min)	
Week 3 2nd Language Comprehension		Listen to the Story (5-15 min) How Words Sound (5 min)		Escuchen Este Cuento (5-15 min) Como Suenan Las Palabras (5 min)
1st Language Production	Dímelo Tu (15 min)	(BSM)*	You Say It (15 min)	(BSM)*
Week 4 Concept Development	PSI Spanish (15 min)		PSI English (15 min)	
Week 5 Make-up	←			→
Week 6 1st Language Comprehension	Escuchen Este Cuento (15 min) Como Suenan Las Palabras (5 min)		Listen to the Story (15 min) How Words Sound (5 min)	
Estimated Time in Minutes	75 to 85 minutes		75 to 85 minutes	

1. It is estimated that there will be a span of 7 months between the pre and post test period.

* Same Administration

Table 8

Summary of Recommended Test Battery

Domain	Test
Language Production	
Descriptive Skills	El Circo--Dímelo Tu/You Say It Part 2 (Button), #2-7 Part 2 (Cinta)
Conversational Development	El Circo--Dímelo Tu/You Say It Part 3 (Narration and Follow-up)
Linguistic Markers	El Circo--Dímelo Tu/You Say It Part I, # 1-27
Language Comprehension	
Word Meanings	El Circo--Escuchen Este Cuento/ Circus Listen to the Story, Items #1,3-12, 14, 16-18, 20-24
Auditory Discrimination	El Circo--Como Suenan Las Palabras/ Circus How Words Sound, all items
Determining the Main Idea	El Circo--Escuchen Este Cuento/ Circus Listen to the Story, Items #1-3, 6, 9-20, 22, 24-25
Concept Development	
Visual Discrimination	Preschool Inventory--Items 5, 6, 25, 28, 29-32, 19-22
Seriation	El Circo--Escuchen Este Cuento/ Circus Listen to the Story, Items #2, 13, 15, 24
Number Concepts *	Preschool Inventory--Items #14-17
Classification/Grouping	El Circo--Escuchen Este Cuento/ Circus Listen to the Story, Items, 2, 13, 15, 24
Spatial Relationships	Preschool Inventory--Items #7, 23, 24, 18-22
Time Relationships	Participant Observation
Symbolic Representation	El Circo--Escuchen Este Cuento/ Circus Listen to the Story, Items, 2, 13, 15, 24
General Knowledge *	Participant Observation
	Preschool Inventory--Items #8-13

Note: * = New Component
Preschool Inventory---Form A, SRI

Test Administration

Once the appropriate tests are chosen, the next important step is the actual collection of the data. As stated earlier, data collection will not start until the Fall of 1979. Pilot testing of the instruments, scheduled from January to March 1979, will help the evaluators develop an adequate plan for test administration, since procedures used to hire, train, and monitor field staff will be evaluated.

Additionally, the recommended battery will be further evaluated for appropriateness, usability, and reliability. Psychometric analysis of the pilot data will include item response distribution, internal consistency, reliability, item-total correlation, whole scale means, and standard deviations.

In relation to the structured observations, we recommend that the child observation systems and the ethnographic observations be further filed-tested during Phase II of the evaluation to determine their usefulness and feasibility. In addition, the above pilot testing will clarify some of the issues listed below.

Time of Administration

This refers to the time it takes to administer the complete test battery. It is our position that a child should not be tested for more than 20-minutes per session and total test time should be kept to a minimum. Collection of the pre- and post-test data should not last more than three weeks in order to minimize the time span between the pre- and the post-test. Both treatment and control groups should have comparable testing times.

Order of Testing

A test sequence has been described above (see Table 7 for a summary) and will be field-tested.

Assembling the Students

This step of the data collection process is quite problematic. Since the design includes a control (no treatment) group, procedures used for pretesting may prove cumbersome. We recommend that the evaluators meet with the staff of each model and key site personnel for careful planning. It is important that all the children, treatment and control, take both pre and post test under similar testing conditions. As a consequence, at each site it is necessary to specify the test locations, the time for each test, make-up sessions, supervision, etc. A few carefully trained administrators will test all the children at each site.

Scoring of the Test

Scoring can usually be done separately from the test admininstration; therefore, for accuracy anf efficiency, scoring will be centralized locally in Juárez and Associates, Los Angeles offices.

Data Analysis

A plan for data analysis is described in the Juárez and Associates revised design and evaluation plan (1978). This data analysis plan will be evaluated once the test selection is finalized. A document that describes the data analysis plan will be reviwed by consultants and by the evaluator's advisory committee.

Appendix A

The Models' Characteristics and Objectives

This document is incomplete and it is included only to illustrate an important step in the identification and selection of objectives for evaluation.

Model A

A. Model Characteristics

1. Cultural relevance and continuity.
2. Language maintenance.
3. Children must be able to handle objects or take part in activity for ideas to form.
4. Growth proceeds as an individual is impelled to engage in more complex ways of being.
5. Must integrate three domains in the curriculum: cognitive-linguistic; socioemotional, and psychomotor.
6. Design learning activities after observations of children, family and community.
7. Children should be encouraged to seek answers w/in available resources.
8. Use scientific method--observing, questioning, exploring, hypothesizing, experimenting, validating--throughout the curriculum.
9. Children must develop skills for categorization, discrimination, generalization, inductive thinking, and making inferences.
10. Not all children will achieve all objectives but will show a balanced achievement.

B. Staff/Adult Language

1. At least 50% of staff be fluent in Spanish or be bilingual.

C. Objectives

1. Socioemotional.
   a. Awareness of self as an individual and as a member of a group.
      1) Children will give name, address, telephone.
      2) Children name Head Start Center, teachers.

Model B

1. Cultural relevance and continuity.
2. Language maintenance.
3. Education about language and culture is experience and cannot necessarily be taught or packaged.
4. Curriculum is not a specific series of lessons or activities; instead they adopt an "open framework" based on knowledge of how children learn and think.
5. Basis of the framework is some 50 "Key Experiences" organized under: active learning, language, experiencing and representing, classification, seriation, number, space, time, sequencing events in time, and bilingual/multi-cultural.
6. Daily routine is consistent involving a child planning, working and reviewing while adult is supporting, encouraging and basically acting as a resource.
7. Children learn from their interactions with people and materials when they have opportunities to explore, predict, observe, describe, and solve problems.

1. Must have a balance of Spanish and English-speaking adults.

1. Miscellaneous.
   a. Child develops a plan.
   b. Child follows his/her plan.
   c. Child solves problems.
   d. Child focuses on activity.

BEST COPY AVAILABLE

Model C

A. Model Characteristics

1. Cultural relevance and continuity.
2. Language and cultural maintenance.
3. Teach children in a manner consonant with their values and experience but provide them with abilities to function in a complex society.
4. Of critical importance in a child's development of communication, human relations, motivation & learning styles are 1) language and heritage; 2) cultural values; 3) teaching styles.
5. Assumes that learning style is a product of 1) a child's style of communicating, relating to others and responding to incentives which result from, 2) a mother's style of teaching and child rearing which in turn are derived from, 3) cultural values and lifestyles.
6. Assures that Spanish speaking cultures emphasize achievement through education and achievement through cooperation. Hence families must be involved educationally and cooperative tasks must be emphasized.
7. Educational programs need to stress close personal relations between teacher and child, be child centered, provide for learning based on modeling, use personalized incentives, provide for structured and directed learning, be humanistic and provide for parent involvement.
8. Bilingual children have an intellectual advantage over monolingual children.
9. Assures the existence of two cognitive styles which must be considered in developing a curriculum.
10. Emphasizes a "Growth Model" for bilingual education involving 1) concurrent use of both languages except for reading and second language instruction for monolinguals, and 2) alternating use of two languages.

Model D

1. Based on an understanding of the children, the culture of the community in the context of societal pluralism, the language spoken, parental involvement and decision-making, keeping costs low and maintaining easy adaptation.
2. Use of two languages; the child's first language and the acquisition of the second.
3. An eclectic blend of theoretical approaches and culturally and linguistically compatible content and methodologies.
4. Integrated goals for children's growth at the cognitive, socio-emotional, psycho-motor and language development levels.
5. Development of the whole child.
6. Claim an eclectic method including the best elements of the permissive-enrichment, structured-environmental, and structured-cognitive approaches.
7. Teachers are seen as change agents and surrogate mothers; nothing happens unless the teacher wants it to happen. Teacher behavior is monitored.

Appendix A (continued)

The Models' Characteristics and Objectives

Model A

3) Children describe themselves physically.

4) Children say how people differ and people and animals differ.

5) Children identify something unique about them, others.

b. Self-Acceptance--capable and worthy.

1) Children eager to pursue the new/complex

2) Children communicate their capability to others.

3) Children evidence pride in accomplishments.

4) Children describe what they like to do.

5) Children show self-confidence through expression.

c. Positive Attitudes toward Own and Other's Cultural Heritage

1) Children describe their own customs.

2) Children discuss customs and dishes.

3) Children describe how a member of a cultural group contributes to society.

4) Children describe an event in the history of a cultural group.

d. Recognition of and Coping with Emotions.

1) Children show and talk about their feelings.

2) Children accept the inappropriateness of certain objects and experiences.

3) Children describe another child's feelings

4) Children practice working cooperatively.

5) Children accept others' shortcomings.

6) Children show trust of others.

e. Independence, Autonomy and Locus of Control

1) Children care for needs without adult assistance.

51

Model B

e. Child creates representations with details using materials.

f. Child recalls something he/she did.

g. Child uses art media to represent something done.

h. Child enjoys being in company of peers.

i. Child shares and takes turns.

j. Child settles arguments verbally without adult help.

k. Child is not vindictive.

l. Child asserts rights in socially acceptable ways.

m. Child talks with others about personally meaningful experiences.

n. Child uses diverse problem-solving strategies.

o. Child does not seek constant praise.

p. Child is active and not distruptive.

q. Child is proud of his efforts.

r. Child attempts to write, trace, copy.

s. Child interested in reading.

t. Child attempts to imitate and repeat words and phrases in appropriate contexts.

u. Develop child's ability to function in "dominant" and native cultures.

v. Develop a positive attitude toward both languages.

w. Interpret and translate for others.

x. Label a language correctly.

y. Interact with others from own and other cultural backgrounds.

z. Express positive affect toward own culture.

BEST COPY AVAILABLE

Model C

B. Staff/Adult Language

1. No mention

C. Objectives

1. Activities center on concept lessons, psychomotor development, language, heritage and culture.

2. Primary objective is to enhance young children's fluency and ability to conceptualize in Spanish and English.

3. Language
   a. Increase vocabulary, both production and comprehension.
   b. Increased variety and appropriateness in language structure.
   c. Ability to explain or narrate events.
   d. Ability to describe
   e. Ability to inquire
   f. Ability to give verbal instructions
   g. Listening and concept comprehension

4. Other Objectives
   a. Enhance a child's self-esteem
   b. Enhance positive attitudes toward both Spanish and English and the respective cultures
   c. Emphasize equally both languages.
   d. Involve parents for interpersonal and intercultural understanding.

5. Other curriculum and Objectives to be developed in Spanish and English.
   a. For ESL, recommend the comercial curriculum "CORE"
   b. Propose to develop a cognitive curriculum that draws heavily from commercial materials with the addition of a bilingual and cultural emphasis.

Model D

1. No mention

1. Developmentally appropriate learning experiences.
   a. Learning experiences occur within the appropriate cultural context.
   b. Develop English and Spanish thru a natural approach to language acquisition.
   c. Children develop basic concepts and skills in their first langu
   d. Develop specific cognitive processes and concepts.
   e. Engage in activities at the operational and perceptual levels appropriate to age group.
   f. Engage in both self-directed and teacher directed learning experiences.
   g. Participation in learning experiences which emphasize socializati and self-concept development.

2. Staff training to individualize instruction, create a safe environmen involve parents, and support the child's cultural identity.
   a. Teachers develop skills in observing, diagnosing and individualizing learning experiences.
   b. Design an appropriate environment.
   c. Bring about optimal parental involvement.
   d. Support child's ethnic identity and self-concept by making cultural experience integral to the program.

3. Optimize parent involvement.
   a. Parents will achieve a greater sense of the importance of their role as teachers of their children

52

Appendix A (continued)

The Models' Characteristics and Objectives

Model A

D. Assessment Instruments.

1. Children Impact Data.

   a. Criterion reference observational checklists being developed. They identify verbal and nonverbal behaviors along with the frequency strength and so on of the behaviors. It may be necessary to wait for certain choice situations in order to observe an appropriate range of behaviors.

2. Teacher Training Impact.

   a. Developing an instrument to assess implementation but may have implications for assessing impact as well.

3. Parent Impact.

   a. None

E. Implementation Evaluation.

1. A "Survey of the Classroom Environment instrument is available and has been pretested. Items are intended to assess the availability and nature of special areas developed for particular curricular activities.

Model B

a. PPLAT (Preschool Production Language Assessment Task). This is a 20 minute tape-recording of one child and one adult interacting over a set of toys and other items. The Spanish version is being developed.

b. Child Observation Record (COR). This is a dichotomously scored listing of statements that describe the child's behavior in the processes and domains deemed important by the program. It generally coincides with the process and key experience variables.

a. PTTP (Preschool Teacher Profile) is an observation checklist scored in gradients under two headings "attainment rating" and "qualifications." The checklist includes items under Room Arrangement, Daily Routine, Work Time, Action, Language, Representation, Classification, Seriation, Number, Space, Time, Feelings, Classroom Management, Team Evaluation and Planning, Working with Parents.

a. Parent Questionnaire is an interview schedule prepared in Spanish and English and includes items related to language use, parent involvement in the classroom and with children and parent organization.

1. The key experiences list, the discussion of the classroom environment in the Marco Abierto document, the Observational Data Sheet and the PTTP might be used in this regard.

BEST COPY AVAILABLE

Model C

D. Assessment Instruments

1. Some interest expressed in a values clarification process in the community and in ethnographic/anecdotal evidence.
2. Indicates will use criterion reference tests.
3. Indicates will use the "Bilingual Syntax Measure"
4. Indicates will use teacher reports for areas of child self-esteem and social skills.
5. Indicates will secure information on frequency and nature of parent involvement as well as use of "home curricular materials" from program records, parent meeting minutes and interviews/survey instruments.
6. Visitation report suggest that staff may be opposed to impact data assessment.
7. Will assess flexibility in the curriculum by relying on Head Start staff viewpoints.

E. Implementation Evaluation

1. Staff perceives this and other forms of evaluation as primarily serving a feedback function.
2. Will include cost of materials training, and other implementation activities.

Model D

1. Children Impact Data (none indicated).
2. Teacher Training - Included is an observation checklist for use by teacher supervisors with the following categories: Assessing children's instructional needs; Classroom Management, Parental Involvement, Interaction Patterns, Classroom Design and Materials, Teaching Methodology (No teacher assessment instruments included).
3. Parent Impact - None indicated although a parent interview questionnaire is available.

1. The program has developed a series of booklets some of which may have relevance here:

   Booklet #1 Concerns the Curriculum Model, #2, #3, #4, #5, #6, #19 Child Growth and Development, #7, #8, #9, #10, #11, #12, #17 Instructional Materials, #13 Teacher-child Interaction, #14 Teacher Interaction, #15, #16 Parent Interviews, #18 Classroom Management, #20 Diagnosing Children, #21 Instructional Design (No mention of Implementation Evaluation).

Appendix B

Cross-Model Behaviors

Domain: Spanish/English Language Production

Component: Oral Language Development - Descriptive Skills

a) ability to identify common objects

(Examples: Classroom items such as playdough, blocks, crayons, paints, etc. Food, clothing, transportation vehicles objects in the community (i.e., signals, street lights, houses, apartments, etc.), plants, animals, furniture, etc.).

b) ability to describe common objects and situations

(Examples: Classroom events (i.e., roll call, lunchtime, naptime, grouptime, storytime, pick-up time, etc.) Daily routines (breakfast, dinner, etc.) special holidays).

c) ability to describe self

(Examples: Body parts, physical dimensions (i.e., tall/short; fat/thin).

d) ability to describe others

(Examples: Body parts, physical dimensions (i.e., tall/short; fat/thin).

e) ability to express feelings

(Examples: Hunger/satiation, happy/sad, hurts/doesn't hurt, sleepy/not sleepy, feels good/doesn't feel good, anger, jealousy).

Component: Oral Language Development - Conversational Development

a) ability to give verbal instructions

(Examples: Directing another on how to perform a specified task, i.e., climbing jungle gym, telling someone where an object is, i.e., under the table).

b) ability to make-up (relate) a story or event; alternative - can make statements about favorite story or event

(Examples: Tells another about a special occurrence, i.e., birthday, special outing, T.V. program).

Component: Oral Language Development, Linguistic Markers

a) ability to use the negative forms in statements

(Examples: Don't, can't, won't, have not, etc.).

b) ability to use interrogative forms by intonation

(Examples: Are you going? Me?).

c) ability to pluralize regular nouns

(Examples: Cup/cups, book/books, shoe/shoes).

d) ability to produce simple sentences

(Examples: I want some juice).

e) ability to produce compound sentences

(Examples: I want some juice and I want some cookies too).

f) ability to produce complex sentences

(Examples: I want some juice but there isn't anymore).

g) ability to use present tenses

(Examples: El conejo brinca).

h) ability to use past tenses

(Examples: El conejo brincó).

Domain: Spanish/English Language Comprehension

Component: Word Meanings

a) identifies spoken words with corresponding pictures

Component: Auditory Discrimination

a) identifies common sounds from the home, school and community

(Examples: Animal noises, street noises).

b) identifies familiar voices

c) distinguishes non-vocal sounds which are alike in a given way

(Examples: Rhythm, intensity, pitch noises).

d) identifies rhyming words

(Examples: True/blue, tool/pool, cambiar/llevar/sentar, lobo/robo).

Component: Determining the Main Idea

a) recall outstanding events from an oral story

(Examples: T.V. program).

b) able to recall what s/he did in the classroom/home

c) able to provide details about classroom/home

Domain: Concept Development

Component: Visual Discrimination

a) identifies likenesses and differences (color, size, shape, and texture) in objects

Component: Seriation/Sequencing

a) arranges objects in sequence

(Examples: Small to large, short to tall, thick to thin).

b) describes the relationship of sequenced items

Component: Classification/Grouping Objects

a) sorts and matches objects (size, form, condition, shape, color, substance)

b) holding more than one attribute in mind at the same time

(Examples: Can you find something that is red and made of wood).

c) distinguishes "some" "all" "more" "less" and "none"

Component: Spatial/Relationship/Space

a) children use terms showing a relationship to time

(Examples: "Mas tarde" "Hasta luego" Times of the day (morning, afternoon, evening) after, before, sometimes etc.).

b) understands "time" unit or ideas

(Examples: Uses time terminology in a way that demonstrates understanding).

c) observes that clocks and calendars are used to mark the passage of time

(Examples: In housekeeping center children note that clocks, watches, and calendars in spontaneous play).

Component: Symbolic Representation

a) uses materials symbolically

(Examples: Uses blocks to make an apartment, house or to represent people, animals or objects. Uses playdough to make "make believe" foods, animals, objects, etc. Uses materials in the community (rocks, leaves, twigs, etc.) to represent items such as cars, people, houses, etc.).

b) makes drawings and paintings

c) imitates actions, sounds

(Examples: Imitates the sound of a fire alarm, imitates the movement of a favorate T.V. personality (The Fonz), made by others (i.e., sharing, bending, turning).

d) pretends

(Examples: Child can assume the role of someone else, i.e., child can take on the role of people, animals (make believe you are (pantomine) you are a bunny rabbit).

Domain: Socio-emotional Development

Component: Socio-emotional Development

a) sustains interest in group activity

(Examples: Child completes a group game such as "Ring Around the Rosy". Child shows interest in group activity for a determined period of time (given contextual restraints)).

b) works cooperatively with others; alternative - child cares for her own needs without continual adult assistance

(Examples: Child appears to enjoy playing with others in a cooperative manner. Given contextual restraints, child gets adult attention in ways that are acceptable for age).

c) follows directions

(Examples: Child makes an effort to follow the social rules (i.e., not hitting others, waiting for turn, etc.).

d) expresses desire to attend the Head Start Program

(Examples: Spontaneous pronoucement).

e) shares and takes turns

(Examples: Child is willing to take turns when the need occurs, allows others to share her materials).

f) evidences pride in accomplishments

(Examples: Indicates happiness or feelings of self-worth over completed project)

g) sees self as capable of mastering new situations

(Examples: Expresses intent to confront new situation/tasks "I can do it.").

h) communicates capabilities to others

(Examples: Reports confidence concerning home/school task).

Appendix C

Sources of Information

Commercial Publishers	Schools/Districts and Universities	Research Institution	Individuals
The Allington Corporation	Brooklyn Bilingual Center for Preschoolers	The Australian Council for Education Research LTD	Mr. Fred Córdova
American Guidance Service	Cincinnati Public Schools	Center for the Study of Evaluation	Ms. Teresa Delgado
Chess and Associates	Institute for Developmental Studies, NYU	Dissemination Center for Bilingual/Bicultural Education	Dr. E.H. Henderson
Consulting Psychological Press	Los Nietos School District	Early Childhood Research	Dr. Martin Kohn
CTB/McGraw-Hill	The Research and Development Center for Teacher Education	Educational Testing Service	Dr. Charles Porterfield
Guidance Testing Associate	Rio Hondo Area Action Council	ERIC Clearinghouse on Tests, Measurement, and Evaluation	Dr. Jane Stallings
Harcourt, Brace and Jovanovich, Inc.	University of California at Los Angeles, Early Childhood and Developmental Studies Department	Rand Corporation	Dr. Rhona Weinstein
Houston Test Company	University of California at Berkeley, Family Socialization and and Developmental Competence Project	Stanford Research Institute	Dr. R. Woolner

Appendix C (Continued)

Sources of Information

Commercial Publishers	Schools/Districts and Universities	Research Institution	Individuals
Instructional Objectives Exchange	Wellesley Public Schools		Dr. E. Zigler
Learning Concepts			
Person-O-Metrics, Inc.			
Priority Innovations, Inc.			
The Psychological Corporation			
Psychometric Affiliates			
Science Research Associates			
SEFA Publications LTD			
South Texas Testing Service			
Southwestern Cooperative Educational Laboratory			
Virginia Research Associates, LTD			
W. Clement and Jessie V. Stone Foundation			
Western Psychological Services			

APPENDIX D

Language Measures

1. Ambiguous Verbal Stimulus Test to Measure Language Dominance in Spanish-English Bilinguals, ETS Test Collections
2. The Appell Test, EDCODYNE Corp.
3. Basic Inventory of Natural Language (BINL), Checpoint Systems
4. Bilingual Scale of Oral Language, Allan Toronto-author
5. Bilingual Syntax Measure, Harcourt, Brace and Jovanovich, Inc.
6. Compresión del Lenguaje Oral, Interamerican Series, Guidance Testing Assoc.
7. Crane Oral Dominance Test, Motivational Learning Programs
8. Dade County Test of Language Development-Receptive, Dade County Public Schools
9. Del Rio Language Screening Test, National Educational Laboratory Publishers Inc.
10. Destrezas de Lenguaje Oral, Bakersfield City School District
11. Diagnostic Language Assessment, Puerto Rican Family Institute
12. Dos Amigos Verbal Language Scales, Academic Therapy Publications
13. El Circo-Como Suenan Las Palabras; Para que Sirven Las Palabras, Dímelo Tú, ETS
14. Examen Exploratorio de Lengua Española, Bilingual Center, City School District #17, N.Y.
15. Flexibility Test to Measure Language Dominance in Spanish-English Bilinguals, ETS Test Collections
16. Fluency Test, Follow Through Project
17. Gloria and David---Oral Bilingual Test, Language Arts, Inc.
18. Home Bilingual Usage Estimate, Rudolph V. Skoczylas
19. Informal Language Assessment, Marysville Joint Unified School District
20. James Language Dominance Test, Learning Concepts
21. Language Assessment Battery (LAB), Houghton Mifflin Company
22. Language Assessment Scales, Linguametrics Group
23. Language Dominance Index Form, California State Department of Education

24. Language Dominance Survey, Los Nietos School District

25. Language Facility Test (LFT), The Allington Corporation

26. Michigan Oral Language Production Tests; Structured Response Test, American Council on the Teaching of Foreign Languages

27. Northwestern Syntax Screening Test, Dr. Laura Lee

28. Oral Language Ability Scale (OLAS), Northeast Center for Curriculum Development

29. Oral Language Evaluation (OLE), Dr. Nicholas J. Silvaroli

30. Oral Language Inventory-Bilingual Education Program, Robstown Independent School District

31. Oral Language Proficiency Scale and ESOL Placement Interview, Dade County Public Schools

32. Oral Language Proficiency Test, Steve Moren

33. OTT Test of Oral Language, ETS Test Collection

34. PAL Oral Language Dominance Measures, El Paso Public Schools

35. The Pictorial Test of Bilingualism and Language Dominance, Texas Testing Service

36. Preschool Language Scale, Charles E. Merrill Publishing Co.

37. Pupil Communication Skills Inventory, Brentwood Union Free School District

38. Pupil Language Usage Inventory, Brooks Consolidated ISD

39. Quick Language Assessment Inventory, Moreno Educational Co.

40. San Bernadino Language Dominance Survey, San Bernadino Bilingual Education Office

41. Screening Test for Auditory Comprehension of Language, Learning Concepts

42. Spanish-English Dominance Assessment Test, ETS Test Collections

43. Spanish/English Language Performance Screening (S/ELPS), CTB/McGraw-Hill

44. Spanish-English Screening Instrument, Bilingual Project-Tucson, Arizona

45. Spanish and English Vocabulary and Concept Comprehension Test, Santa Clara County Offices of Education

46. Spanish Test of Oral Proficiency, Bay Area Bilingual Education League (BABEL)

47. SWCEL Comparative Language Dominance Test, Southwestern Cooperative Educational Laboratory, Inc.

48. SWCEL Spanish Oral Capacity Test, Texas Education Agency

49. Test for Auditory Comprehension of Language (TACL), Learning Concepts, Inc.

50. Test of Basic Language Comprehension in English and Spanish Form A, E.J. Cervenka

51. Test of English Grammar, Santa Clara County Office of Education

52. Test of Language Dominance (TOLD) Spanish Version, Southwest Research Associates, Inc.

53. Test of Oral English Production, Robert H. Landen

54. Tucson Public School Language Assessment, Tucson Public Schools

LANGUAGE COMPREHENSION

Name of Test _____ Subtest _____

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Word Meanings a) identifies spoken words with corresponding pictures Additional Concepts b) Identifies spoken phrases, sentences, story with pictures				
Component Auditory Discrimination a) identifies common sounds from home and community b) identifies familiar voices c) distinguishes non-vocal sounds which are alike in a given way d) supplies words with the same initial sound e) identifies rhyming words Additional Concepts				

LANGUAGE COMPREHENSION

Name of Test _____ Subtest _____

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Determining the Main Idea a) recall outstanding events from an oral story (short-term) b) recall outstanding events from an oral story (long-term) c) recalls what s/he did in the classroom/home d) provides details about the classroom/home Additional Concepts				

*Legend

Judged Degree of Match:

0 = poor
1 = acceptable
2 = very good

Judged Degree of Item Difficulty:

0 = very easy or very difficult
1 = acceptable
2 = very good

LANGUAGE PRODUCTION

Name of Test _____ Subtest _____

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Oral Language Development Descriptive Skills a) ability to identify common objects b) ability to describe common objects and situations c) ability to describe self d) ability to describe others e) ability to express feelings Additional Concepts				
Component Oral Language Development Conversational Development a) ability to give verbal instructions b) ability to make-up (relate) a story or event; alternative-can make statements about favorite story or event Additional Concepts				

LANGUAGE PRODUCTION

Name of Test _____ Subtests _____

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Oral Language Development Linguistic Markers				
a) ability to use the negative form in statements				
b) ability to use interrogative forms by intonation				
c) ability to pluralize regular nouns				
d) ability to produce sentences				
e) ability to produce compound sentences				
f) ability to produce complex sentences				
g) ability to use present tenses				
h) ability to use past tenses				
Additional Concepts				
i) ability to use appropriate verb forms				
j) ability to use appropriate adverbs				
k) ability to use possessives				

*Legend

Judged Degree of Match:

0 = poor
1 = acceptable
2 = very good

Judged Degree of Item Difficulty:

0 = very easy or very difficult
1 = acceptable
2 = very good

CONCEPT DEVELOPMENT

Name of Test _____ Subtest _____

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Visual Discrimination a) identified likenesses and differences (color, size, shape and texture) in objects Additional Concepts				
Component Seriation/Sequencing a) arranges objects in sequence, matches objects, letters, and numbers in sequence b) describes the relationship of sequenced items c) identify correct sequence of numbers Additional Concepts				
Component Number Concepts a) counting b) matching sets (one/one) c) matching numbers to sets d) numerical terms				

CONCEPT DEVELOPMENT

Name of Test _____ Subtest _____

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Matching/Classification/ Grouping of Objects a) sorts and matches objects (size, form, condition, shape, color and substance) (visual matching) b) holding more than one attribute in mind at the same time c) distinguishes "some" "all", "more", "less" and "none" Additional Concepts d) relational terms				
Component Spatial Relationship/Space a) demonstrates and describes relative position of things Additional Concepts				
Component Time Relationships a) uses terms showing a relationship to time				

CONCEPT DEVELOPMENT

Name of Test _____ Subtest _____

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
b) understands "time" units or idea c) observes that clocks and calendars are used to mark the passage of time Additional Concepts				
Component Symbolic Representation a) uses materials sumbolically b) makes drawings and paintings c) imitates actions, sounds d) pretends Additional Concepts				

*Legend

Judged Degree of Match:
0 = poor
1 = acceptable
2 = very good

Judged Degree of Item Difficulty:
0 = very easy or very difficult
1 = acceptable
2 = very good

APPENDIX F

LANGUAGE PRODUCTION

Name of Test Bilingual Syntax Measure (Language Dominance)

Subtests

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Oral Language Development Linguistic Markers				
a) ability to use the negative form in statements				
b) ability to use interrogative forms by intonation				
c) ability to pluralize regular nouns	8	1	1	
d) ability to produce sentences				
e) ability to produce compound sentences				
f) ability to produce complex sentences				
g) ability to use present tenses				
h) ability to use past tenses				
Additional Concepts				
i) ability to use appropriate verb forms				
j) ability to use appropriate adverbs				
k) ability to use possessives	18, 20	2	2	

*Legend

Judged Degree of Match:
0 = poor
1 = acceptable
2 = very good

Judged Degree of Item Difficulty:
0 = very easy or very difficult
1 = acceptable
2 = very good

LANGUAGE PRODUCTION

Name of Test Bilingual Syntax Measure (Language Dominance) Subtest

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Oral Language Development Descriptive Skills				Recommended
a) ability to identify common objects /things	1, 6, 7, 8	2	2	
b) ability to describe common objects and situations	16, 17	2	2	
c) ability to describe self				
d) ability to describe others				
e) ability to express feelings				
Additional Concepts				
Ability to make conceptual inference	3, 4, 5, 9, 11	1	2	
Component Oral Language Development Conversational Development				
a) ability to give verbal instructions	2			
b) ability to make-up (relate) a story or event; alternative-	13, 22, 23, 24	2	2	
can make statements about favorite story or event	14, 15			
Additional Concepts				

LANGUAGE PRODUCTION

Name of Test EL CIRCO Subtest 10S Dímelo Tú/You Say It

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Oral Language Development Descriptive Skills				Recommended
a) ability to identify common objects	10S (II, #1)	1	0	
b) ability to describe common objects and situations	10S Part Two Button and Shoelace	2	2	
c) ability to describe self				
d) ability to describe others				
e) ability to express feelings				
Additional Concepts				
Component Oral Language Development Conversational Development				
a) ability to give verbal instructions				
b) ability to make-up (relate) a story or event; alternative-can make statements about favorite story or event	10S Part Three Narration & Follow-Up Questions		2	
Additional Concepts				

LANGUAGE PRODUCTION

Name of Test EL CIRCO

Subtests 10S Dímelo Tú/You Say It

Cross-Modal Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Oral Language Development Linguistic Markers				
a) ability to use the negative form in statements				
b) ability to use interrogative forms by intonation				
c) ability to pluralize regular nouns				
d) ability to produce sentences				
e) ability to produce compound sentences				
f) ability to produce complex sentences				
g) ability to use present tenses	10S (1-6)	2	2	
h) ability to use past tenses	10S (7-10)	2	2	
Additional Concepts				
i) ability to use appropriate verb forms	10S (11-13, 24, 25)	2	2	
j) ability to use appropriate adverbs	10S (14-19)	2	2	
k) ability to use possessives	10S(20-23, 26, 27)	2	2	

*Legend

Judged Degree of Match:

0 = poor
1 = acceptable
2 = very good

Judged Degree of Item Difficulty:

0 = very easy or very difficult
1 = acceptable
2 = very good

LANGUAGE COMPREHENSION

Name of Test El CIRCO

Subtest #7 Como Suenan Las Palabras/How Words Sound

#9 Escuchen Este Cuento/Listen to the Story

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Word Meanings				Recommended
a) identifies spoken words with corresponding pictures				
Additional Concepts				
b) Identifies spoken phrases, sentences, story with pictures	# 9 (1, 3-13,14 16-18, 20-24)	2	2	
Component Auditory Discrimination				
a) identifies common sounds from home and community				
b) identifies familiar voices				
c) distinguishes non-vocal sounds which are alike in a given way				
d) supplies words with the same initial sound				
e) identifies rhyming words				
Additional Concepts				
f) Discriminating Phonemes	# 7 all items	2	2	

LANGUAGE COMPREHENSION

Name of Test EL CIRCO

Subtest # 7 Como Suenan Las Palabras
9 Escuchen Este Cuento

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Determining the Main Idea				
a) recall outstanding events from an oral story (short-term)	#9 (1-3, 6, 9-20 22, 24, 25)	1	2	
b) recall outstanding events from an oral story (long-term)	#9 (24)	1	2	
c) recalls what s/he did in the classroom/home				
d) provides details about the classroom/home				
Additional Concepts				

*Legend

Judged Degree of Match:
0 = poor
1 = acceptable
2 = very good

Judged Degree of Item Difficulty:
0 = very easy or very difficult
1 = acceptable
2 = very good

CONCEPT DEVELOPMENT

Name of Test Cooperative Preschool Inventory Subtest # 9 Escuchen Este Cuento

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Matching/Classification/ Grouping of Objects				
a) sorts and matches objects (size, form, condition, shape, color and substance) (visual matching)	7	1	1	
b) holding more than one attribute in mind at the same time				
c) distinguishes "some" "all", "more", "less" and "none"	23, 24	2	1	
Additional Concepts				
d) relational terms	5, 7, 18-22, 2, 13, 15, 24(9)	1	2	
Component Spatial Relationship/Space				
a) demonstrates and describes relative position of things				
Additional Concepts				
Component Time Relationships				
a) uses terms showing a relationship to time	2, 13, 15, 24(9)	1	1	

LANGUAGE PRODUCTION/CONCEPT DEVELOPMENT

Name of Test Cooperative Preschool Inventory Subtest #9 Escuchen Este Cuento

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Oral Language Development Descriptive Skills a) ability to identify common objects b) ability to describe common objects and situations c) ability to describe self d) ability to describe others e) ability to express feelings Additional Concepts f) Identifies self (personal awareness and body parts)	1-4	2	2	
Component Oral Language Development Conversational Development a) ability to give verbal instructions b) ability to make-up (relate) a story or event; alternative-can make statements about favorite story or event Additional Concepts				

LANGUAGE COMPREHENSION/CONCEPT DEVELOPMENT

Name of Test Cooperative Preschool Inventory Subtest #9 Escuchen Este Cuento

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Determining the Main Idea				
a) recall outstanding events from an oral story (short-term)				
b) recall outstanding events from an oral story (long-term)				
c) recalls what s/he did in the classroom/home				
d) provides details about the classroom/home				
Additional Concepts				
e) makes/understands conceptual inference	13, 15, 16, 18-21, 24, 25 (9)			
Additional Component General Knowledge	8-13	1	2	

*Legend

Judged Degree of Match:

0 = poor
1 = acceptable
2 = very good

Judged Degree of Item Difficulty:

0 = very easy or very difficult
1 = acceptable
2 = very good

CONCEPT DEVELOPMENT

Name of Test Cooperative Preschool Inventory Subtest # 9 Escuchen Este Cuento

Cross-Model Objectives	List of Test Items*	Judged Degree of Match*	Judged Degree of Item Difficulty*	Comments
Component Visual Discrimination a) identified likenesses and differences (color, size, shape and texture) in objects	5, 6, 25, 28	2	2	Psychomotor Dev Items # 26, 27 Recommended
Additional Concepts b) Identifies attributes or properties of an object	29 -32	2	1	
Component Seriation/Sequencing a) arranges objects in sequence, matches objects, letters, and numbers in sequence				
b) describes the relationship of sequenced items	19-22 2, 13, 15, 24 (9)	2	2	
c) identify correct sequence of numbers				
Additional Concepts				
Component Number Concepts				
a) counting				
b) matching sets (one/one)				
c) matching numbers to sets				
d) numerical terms	14-17	2	2	

Appendix G

Instruments Reviewed but not Recommended as Child Impact Measures

Language Tests

- BASIC INVENTORY OF NATURAL LANGUAGE (BINL)

  This is a relatively new test which purports to give a dominance estimate based on language production. The test has good face validity, but was rejected because of limited data on preschool subjects, absence of information regarding use in evaluations and a cumbersome scoring procedure which involves sending protocols to the test maker for computer scoring.

- SCREENING TEST OF AUDITORY COMPREHENSION OF LANGUAGE (CARROW)

  This test has been used widely in a number of evaluation efforts to determine language dominance. However, the items in the test, when used with preschool children, are strongly confounded with school readiness. For this reason it was judged that this test could not yield a "true" measure of dominance in 3 and 4 year old children.

- DEL RIO LANGUAGE SCREENING TEST

  The Del Rio Test is designed to identify difficulties in the child's Spanish and English language abilities. While meeting a number of test analysis criteria, this test was developed and standardized on a sample from rural Southern Texas. Many items were, therefore, judged to have a strong cultural bias for that area. Additionally, there were no data available on its use in nationwide efforts or its applicability in different geographic areas.

- JAMES LANGUAGE DOMINANCE TEST

  The James Test was originally considered a strong contender for inclusion in the test battery because of its brevity and straightforward administration. It was noted that English speaking 3 through 5 year olds were utilized in the development, however, the youngest sample of Spanish speakers was kindergarteners. It was therefore considered inappropriate for use with Spanish speaking preschoolers.

- PICTORIAL TEST OF BILINGUALISM AND LANGUAGE DOMINANCE

  This test includes both comprehension and production components. Its limitations include a narrow sampling of linguistic forms (only nouns) on which dominance is based, and the development and standardization was limited to a South Texas Mexican-American population. In addition, inadequate information was provided regarding its viability for a nationwide evaluation.

Appendix G con't

- TOBE 2

  Owing to the fact that the technical manual for the TOBE 2 has not yet been completed, much technical information on the test was not available for review. However, this test (3 subtests) was rejected on the basis of the inappropriateness of various items for the target population as well as the high level of difficulty for preschoolers. In addition, the TOBE 2 exhibited inadequate congruence to the curriculum models.

- WALKER READINESS TEST

  Although the Walker Readiness Test has been used widely with Head Start evaluations and is relatively straightforward in content and administration, a number of problems were noted with this test. Principally, the range of behaviors sampled is extremely narrow. This being the case, the match between the Walker Test and the various curriculums is not a strong one. Additionally, sample items are sometimes confusing and do not follow the actual test item descriptions. It also should be noted that the Spanish translation is poor.

Appendix G con't

- SWCEL COMPARATIVE LANGUAGE DOMINANCE TEST

This test is a diagnostic instrument which involves a breadth of assessment that is unnecessarily lengthy for a program evaluation such as the one being undertaken. Testing time and training requirements for administration of this test are excessively long and costly.

- TEST OF BASIC COMPETENCE IN ENGLISH AND SPANISH (CERVENKA)

Although the overall review of this test was postive, it has many limitation for use in an evaluation effort. The method for arriving at a dominance score is cumbersome and lacks refinement. Also, items were developed utilizing Mexican-American children and technical data on the instrument was not available for review. It was judged that a number of modifications would be required in order to make this test appropriate for use in a nationwide evaluation.

Concept Development

- CIRCUS/CIRCO SUBTESTS:

PARECIDOS/LOOK ALIKES; MIRA Y RECUERDA/SEE AND REMEMBER
CUANTO Y CUANTOS/HOW MUCH AND HOW MANY

Test items are very similar for each pair of these subtests. In general, these subtests were considered to be well constructed, straightforward and graphically adequate. However, the time requirments for administration were considered excessive for the narrow sample of behaviors each measured. That is to say, that in order to get a profile of concept development it would be necessary to use these tests in combination, thus requiring a lengthy testing period.

PENSANDOLO BIEN/THINK IT THROUGH

As with the above mentioned Circo/Circus subtests, item content between these two concept development tests are substantially similar. Although Pensándolo Bien/ Think it Through have considerable face validity, several questions were identified that would have made utilization problematic. For example, item difficulty on a number of items was judged to be inappropriate for the groups to be tested. Graphic illustrations were occasionally confusing and judged too small and detailed for preschool children.

Appendix H

DÍMELO TU/YOU SAY IT

Publisher/Author: Educational Testing Service, Princeton, New Jersey
Addison-Wesley, San Francisco, California

Content: Language Production

Description: Assesses functional language. Divided into three sub-sections which are scored separately: functional language, description and narration.

Age Range: Preschool and kindergarten (4, 5, and 6 years)

Administration:

- Test time: approximately 15 minutes
- Group/individual: individual
- Who administers: Specialist not required, minimum training required, must be bilingual
- How administered: Oral language elicited through questions, pictures and objects.
- Response mode: Oral

Comments: English and Spanish version has been developed for complementary use.

Validity: High face validity. Other validities not reported.

Reliability: Dímelo Tú: internal reliability: .71-.79, parts I & III, .25-.48, part II
You Say It: internal reliabilities not available at this writing

Norms: 2,034 pre-K children used in the initial piloting and standardization of the instruments for the Circo battery, this included Puerto-Rican, Cuban American and Mexican-American subjects.

BILINGUAL SYNTAX MEASURE

Publisher/Author: Marina Burt, Heidi Dulay and Eduardo Hernandez
Harcourt, Brace and Jovanovich, Inc,

Content: Language Dominance and Production

Description: This test is designed to measure children's oral proficiency in English and/or Spanish grammatical structures by using natural speech as a basis for making judgments. Simple questions are used with cartoon-type colored pictures to provide the framework for a conversation with the child. A comparison of proficiency levels in 2 languages yields information that indicates the dominant languages and degree of bilingual proficiency.

Age Range: 4 years through 9 years

Administration:

Test Time:	10-15 minutes
Group/Individual:	individual
Who administers:	No training necessary (must be bilingual)
How administered:	Interview
Response mode:	Oral
Comments:	This test can be used as a test of language dominance as well as a test of the degree of maintenance or loss of certain basic Spanish grammatical structures in children who acquire Spanish.

Validity: High face validity. The Rationale and Technical Report was not available for review at the time of this writing.

Norms: Norms on 2,000 English speaking children and 2,000 Spanish speaking children tested in March, 1974 are available from the publishers.

ESCUCHEN ESTE CUENTO/LISTEN TO THE STORY

Publisher/Author: Educational Testing Service, Princeton, New Jersey
Addison-Wesley, San Francisco, California

Content: Language comprehension

Description: Consists of 25 items and taps comprehension, interpretation and recall of oral language.

Age Range: Preschool and kindergarten

Administration:

Test time: approximately 15 minutes

Group/individual: individual testing recommended with children four and under

Who administers: minimum training necessary; must be bilingual

How administered: test is read to the child

Response mode: pencil/paper (mark the correct picture)

Validity: High face validity. Appropriate to age level.
Evidence of response to the effects of schooling.

Reliability: Escuchen este Cuento: not available at time of writing
Listen to the Story: .77 total alpha reliability (internal consistency)

Norms: Escuchen este Cuento: 2,034 pre-k included in the initial piloting and standardization of the Circo battery. All major Spanish speaking groups and geographical locations included.
Listen to the Story: 269 nursery school, four major geographical areas represented. Sample biased in favor of middle and upper class.

COMO SUENAN LAS PALABRAS/HOW WORDS SOUND

Publisher/Author: Educational Testing Service, Princeton, New Jersey
Addison-Wesley, San Francisco, California

Content: Language comprehension

Description: English version consists of 44, Spanish version 29 items. Both tap auditory discrimination of phonemes including intial and final consonants and medial vowels.

Age Range: four to six years

Administration:

Test time: approximately 5 to 10 minutes
Group/individual: recommended individual with four year old children
Who administers: minimum training required; must be bilingual and familiar with the spoken Spanish of the local area
How administered: test is read to child
Response mode: paper/pencil (mark the correct picture)

Validity: Good face validity. Appropriate to age level.

Reliability: Como suenan las palabras: alpha reliabilities not yet computed and available at time of writing.
How words sound: .92 internal consistency

Norms: Como Suenan las palabras: 2,034 children represented all major Spanish speaking groups utilized in the initial piloting and standardization of the Circo battery.
How words sound: 300 pre-K children. Middle class bias.

PRESCHOOL INVENTORY-SRI 32 ITEM VERSION

Publisher/Author: Addison-Wesley, San Francisco, California
Stanford Research Institute, Palo Alto, California

Content: Language development, general cognitive skills and concepts

Description: Children are required to name parts of the body, copy designs and label objects. Abilities assessed include: speaking, comprehension, following directions, and quantification.

Age Range: Three to six years

Administration:

Test time: Approximately 15 minutes

Group/individual: individual recommended for children four and under

Who administers: Specialist not required; must be fluent in the language of the test (Spanish or English)

How administered: Examiner reads items to the child, some utilization of props, paper/pencil/crayon materials.

Response mode: Oral, physical manipulation of materials, motoric and paper/pencil.

Comments: 32 item version has been used in several nationwide evaulations and has >.90 correlation with the 64 item ETS version.

Validity: High concurrent validity with a number of other major tests. Established predictive validity.

Reliability: Test-retest: .83-.86; internal consistency: .83-.85

PSI - continued

Norms:

English 64 item: 1970 report on 1531 children from a nationwide sample.
Spanish 64 item: 970 children from all major Spanish speaking geographical centers.

ERRATUM

P. 13, Table 3, English Language Comprehension, SWCEL Test was inadvertently omitted from the table. Analysis is approximately the same as noted for Spanish Language Comprehension.

REFERENCES

American Institutes for Research, Evaluation of the Impact of ESEA Title VII Spanish/English Education Program, Vol. 1: Study Design and Interim Findings, 1977.

Coller, A.R. The Assessment of Self-Concept in Early Childhood Education. ERIC Clearinghouse on Early Childhood Education, 1971.

Cordova, F. El Circo: Test Development for young 4-, 5-, and 6-year old Spanish-speaking children in the United States--An Overview. Paper presented to the Rocky Mountain Educational Research Association, El Paso, Texas, November 1977.

Deloria, D., Love, J., Goedinghaus, L., Gordon, S., Hanvey, R., Hockman, E., Platt, J., & Nauta, M. The National Home Start Evaluation: Summative evaluation results (Interim Report IV). Ypsilanti, Mich.: High/Scope Educational Research Foundation, 1974.

Dotseth, M., Hunter, R., & Walker, C.B. Survey of test selectors' concerns and the test selection process. CSE Report #107, Center for the Study of Evaluation, Los Angeles, 1978.

Hoepfner, R., et al. CSE Secondary School Test Evaluations. Los Angeles: Center for the Study of Evaluation, 1974.

Hoepfner, R., et al. CSE Elementary School Test Evaluations. Los Angeles: Center for the Study of Evaluation, 1976.

Irizarry, R. Personal communication. Center for the Study of Evaluation, U.C.L.A., 1978.

Juárez and Associates, Inc. Revised Evaluation Design and Field Plan. An Evaluation of the Head Start Bilingual Bicultural Curriculum Development Projects. Report prepared for the Administration for Children, Youth, and Families, Contract No. HEW-105-77-1048, April, 1978.

Love, J., Wacker, S., Meece, J. Recommendations for Measuring Program Impact. Interim Report II, Part B. A Process Evaluation of Project Developmental Continuity, High Scope Educational Research Foundation, 1975.

Moll, L.C., et al., "Evaluation of Bilingual/Bicultural Head Start Curriculum Development Project. The Initial Phase: A Cooperative Effort by Educators and Evaluators." Proceedings of the San Antonio Bilingual Head Start Conference, San Antonio, Texas, November 15, 1977.

Kaizen, S., and Bobrow, S.B. Design for a national evaluation of social competence in Head Start Children. Santa Monica, Calif.: Rand Corporation.

Shipman, V. "Disadvantaged Children and Their First School Experiences" ETS - Head Start Longitudinal Study, (PR-72-27) Princeton, New Jersey, December 1972.

Stallings, J., Wilcox, M., Travers, J. "Phase II Instruments for the National Day Care Cost-Effects Study: Instrument Selection and Field Testing" Stanford Research Institute, Menlo Park, California, January 1976.

Walker, C. Standards for Evaluating Criterion-Referenced Tests. CSE Report, 1978.

Walker, D.K. Socioemotional Measures for Preschool and Kindergarten Children, Jossey-Bass, San Francisco, California 1973.